

# PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://spiedigitallibrary.org/conference-proceedings-of-spie)

## High performance astronomical data communications in the LSST data management system

Jeff Kantor, Ron Lambert, Chip Cox, Deborah Levine, Chris Smith, et al.

Jeff Kantor, Ron Lambert, Chip Cox, Deborah Levine, Chris Smith, Arun Jagatheeson, Chris Cribbs, "High performance astronomical data communications in the LSST data management system," Proc. SPIE 7019, Advanced Software and Control for Astronomy II, 701919 (15 July 2008); doi: 10.1117/12.790194

**SPIE.**

Event: SPIE Astronomical Telescopes + Instrumentation, 2008, Marseille, France

# High performance astronomical data communications in the LSST Data Management System

Jeff Kantor, LSST Corporation [jkantor@lsst.org](mailto:jkantor@lsst.org)

Ron Lambert, LSST Corporation [rlambert@lsst.org](mailto:rlambert@lsst.org)

Chip Cox, WHREN-LILA [chipcox@mac.com](mailto:chipcox@mac.com)

Deborah Levine, IPAC/Caltech [deblev@ipac.caltech.edu](mailto:deblev@ipac.caltech.edu)

Chris Smith, LSST Corporation [csmith@ctio.noao.edu](mailto:csmith@ctio.noao.edu)

Arun Jagatheeson, LSST Corporation [arun@sdsc.edu](mailto:arun@sdsc.edu)

Chris Cribbs, LSST Corporation [ccribbs@ncsa.uiuc.edu](mailto:ccribbs@ncsa.uiuc.edu)

## ABSTRACT

The Large Synoptic Survey Telescope (LSST) is an 8.4m (6.5m effective), wide-field (9.6 degree<sup>2</sup>), ground-based telescope with a 3.2 GPixel camera. It will survey over 20,000 degree<sup>2</sup> with 1,000 re-visits over 10 years in six visible bands, and is scheduled to begin full scientific operations in 2016. The Data Management System will acquire and process the images, issue transient alerts, and catalog the world's largest database of optical astronomical data. Every 24 hours, 15 terabytes of raw data will be transferred via redundant 10 Gbps fiber optics down from the mountain summit at Cerro Pachon, Chile to the Base Facility in La Serena for transient alert processing. Simultaneously, the data will be transferred at 2.5Gbps over fiber optics to the Archive Center in Champaign, Illinois for archiving and further scientific processing and creation of scientific data catalogs. Finally, the Archive Center will distribute the processed data and catalogs at 10Gbps to a number Data Access Centers for scientific ,educational, and public access. Redundant storage and network bandwidth is built into the design of the system. The current networking acquisition strategy involves leveraging existing dark fiber to handle within Chile, Chile – U.S. and within U.S. links. There are a significant number of carriers and networks involved and coordinating the acquisition, deployment, and operations of this capability. Advanced protocols are being investigated during our Research and Development phase to address anticipated challenges in effective utilization. We describe the data communications requirements, architecture, and acquisition strategy in this paper.

## Large Synoptic Survey Telescope (LSST) Overview

The LSST will be located on Cerro Pachon, Chile and will achieve full scientific operations in 2015. It is designed to image the whole Southern sky every few nights for 10 years, producing a movie like window into the dynamic Universe.

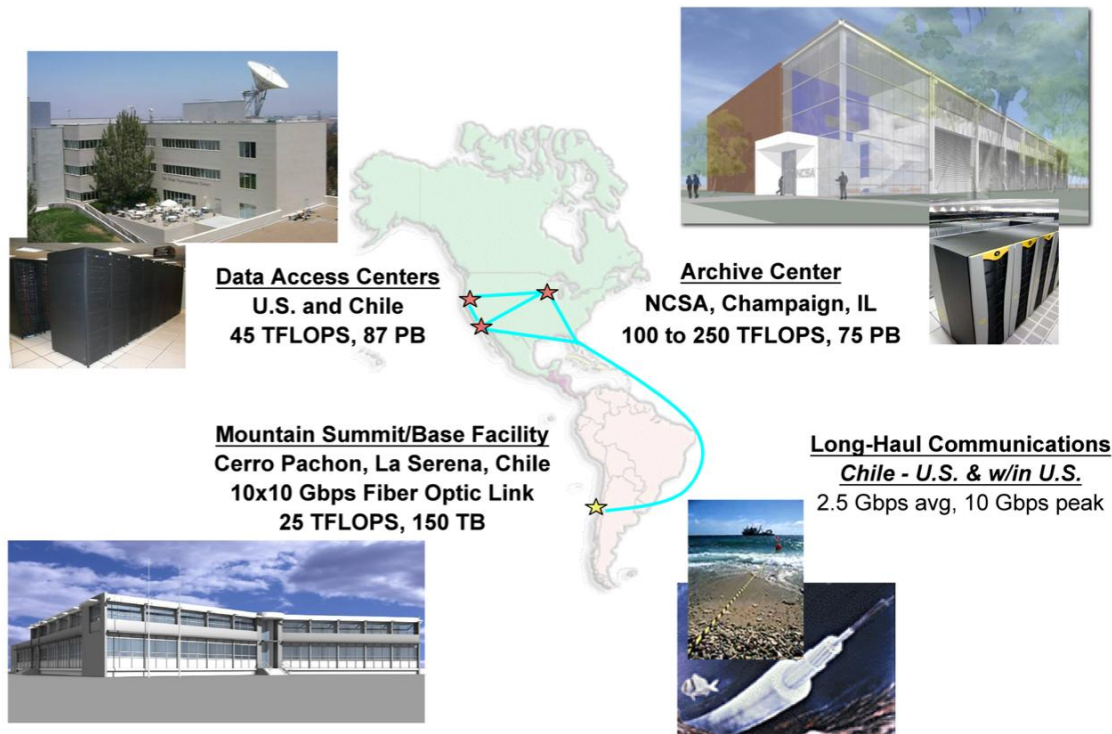


LSST's fundamental design characteristics are:

- 8.4 M Primary Aperture
- 3.5 Degree Field Of View
- 6 Filter Bands (ugrizY)
- 3.2 Billion Pixel Camera
- ~40 Second Cadence, 2 Second Readout
- Two 15 second exposures per pointing
- Public Data (no proprietary period), including
  - Alerts of new events
  - Catalogs of object
  - Archives of images

## LSST Data Management System

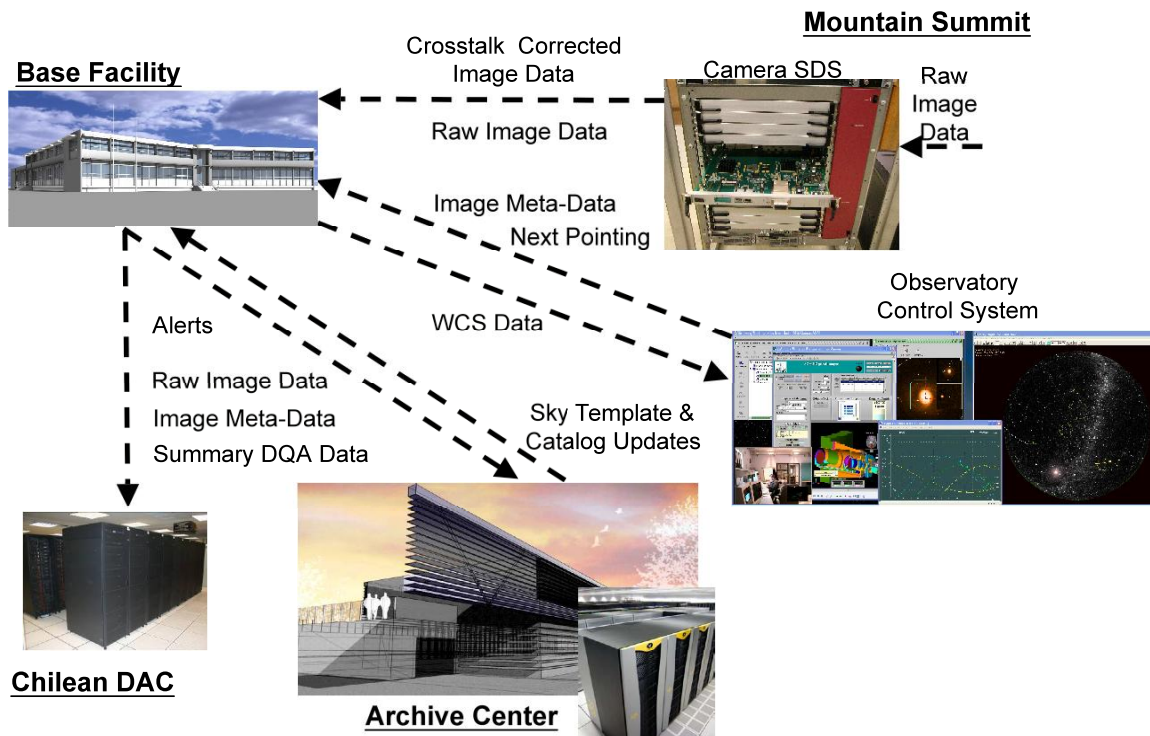
The LSST Data Management System (DMS) is composed of 3 major facility types, inter-connected via high-speed networks.



The Mountain Summit site is located on Cerro Pachon, Chile at the telescope site, and performs readout of the camera and crosstalk correction of raw image data. The Base Facility is located at the Cerro-Tololo InterAmerican Observatory (CTIO) in La Serena, Chile and provides processing and storage for transient alerting. The Archive Center is located at the University of Illinois National Center for Supercomputing Applications (NCSA) in Champaign, Illinois. It provides complete data archiving and computing for re-processing and data release. The Data Access Centers (DAC) provide data replication and computing and storage resources for end users of LSST data. One DAC will be co-located with the Base Facility in Chile. Other DACs will be located in the U.S. and possibly other continents.

## LSST Data Flows and Bandwidth Requirements

Prior to the start of nightly observing, the latest sky template image and astronomical catalog updates are transferred from the Archive Center in Champaign Illinois to the Base Facility in La Serena, Chile. With each pointing, the Observatory Control System sends the next pointing to the Base Facility for moving object predictions. On camera readout and crosstalk correction at the mountain summit, the crosstalk-corrected images flow via the Mountain-Base Network to the Base Facility.



There, the image is processed to remove the instrumental signature, calibrate it, and subtract it with a sky template to detect difference sources. The resultant World Coordinate System (WCS) is sent up to the summit for use as a quality check by the OCS. For every pair of images in a visit, the difference sources are associated with known objects and moving object position predictions. Transient events are detected and alerts are generated.

After the alerts are generated, the raw images are downloaded from the summit to the Base Facility. The alerts and raw images are transferred to the Archive Center and from there on to the DACs. One exception is the Chilean DAC, which receives the raw images directly from the Base Facility to eliminate unnecessary network hops.

From the Archive Center, the alerts are sent to alert distributors such as eStar, GCN, and VOEventNet. After the nightly observing is done, Summary Data Quality Analysis statistics are generated at the Base Facility and sent to the Archive Center.

The OCS' Engineering and Facility Database containing all metadata and telemetry is downloaded to the Base Facility during daylight hours, and forwarded to the Archive Center for permanent archiving. Before the next night's observing starts, the entire night's raw images have arrived at the Archive and are processed again and the object catalogs are updated.

Every 12 months, the Archive also re-processes the entire survey to date, in order to globally calibrate it, and to create a Data Release. This release is sent to LSST Science Centers (not part of the DMS) for scientific validation. On validation, the Data Release is sent to the DACs.

## Mountain Summit Infrastructure

The summit computer room will reside at the summit on Cerro Pachon, which is at 9000 feet elevation. There are three main elements at the summit: the Camera control system (CCS), the Telescope control system (TCS) and the Data Management System (DMS). Each of these systems will be developed independently and rely on the network infrastructure to connect them. The network facilitates data flow between the elements and transfers images and other data to and from the Base Facility.

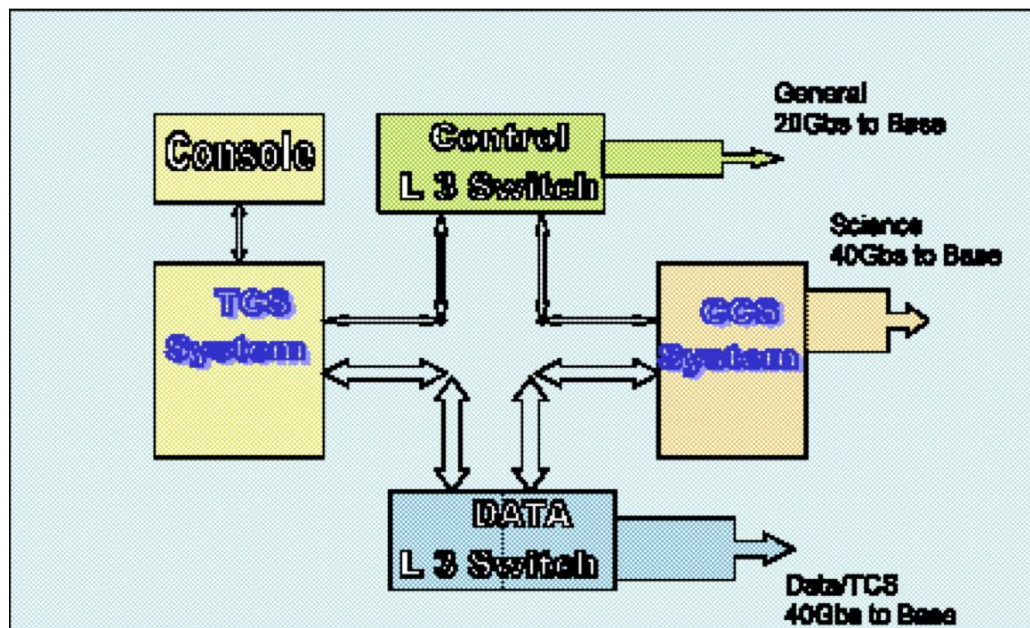
At an early stage Infiniband was considered as a transport but due to the immaturity of Infiniband technology over this distance and lack of market penetration, along with the emergence of 10Gb Ethernet, the latter was considered more



appropriate. The network must be redundant and have failover functionality wherever possible in order to achieve required reliability, thus a major network manufacturer will be selected for providing the equipment.

Ideally we should adopt a mesh architecture with Access and Core level switch/routers, with dual trunk *10 Gigabit* transfer rates between the Access and Core and *1 Gigabit* at the Access level to the individual VLANs. The exception to that is for image data exiting from the Camera system readout destined for the Base Facility, which will be *4X10 Gigabit* ports. These data will not enter the general mountain network as the policy is to be available at the Base Facility in close to real time. The Camera system may also output a second complete image for real time display and other monitoring and diagnostic operations. The CCS will store two days worth of raw image data in solid state memory in the event that there is a disruption in the link to the Base Facility and for any queries of recent data.

The summit will not incorporate a firewall as it is in the trusted network with private connection to the Base. However we may utilize packet-sniffing techniques in the level 3 switches.



As indicated in the diagram Control and Data lines will run over separate networks in order to decrease latency in the system. We will maintain separate subnet/VLANs between the different entities in order to impose levels of security with access control lists.

## Summit to Base Network

The distance from the summit to the base is 55Km point to point or 90Km via the public highway and Observatory access roads. We will leverage current installed public and private Telegraph poles from the Summit to the Base, to install an aerial Corning LEAF fiber bundle with 20 light streams for active duplex circuits, and dark streams for redundancy. This will be a private cable maintained and operated by LSST. The plan below shows the routing.



The bandwidth that will be lit on this bundle is *100Gbs*, via 10 streams of *10Gbs*. The transceivers selected are *10Gbs* Long range XR Xenpaks that have a total power budget of 28dbm. It is difficult to calculate the attenuation of the cable before it is actually installed but we believe it will be on the order of -22dbm, which allows for overhead. The stated policy is that we will NOT utilize a repeater along the circuit so in the event that we do not achieve the attenuation as calculated then our fallback position is a DWDM installation whose attenuation budget gives 34dbm. This would not be our first selection, as DWDM requires a higher level of maintenance and cost.

We believe that should a break occur we can locate and completely repair the LSST fibers within a 24 hour period, with 8 hours required for restoring basic connectivity. The cable will be delivered in 12Km spools, which we will fuse to form the required length.

The 100Gbs total bandwidth will be allocated as follows:

- 40 Gbps Science stream for the real time image data
- 40 Gbps Data stream for secondary image for health display or other usage
- 10 Gbps TCS/OCS stream for telescope control data and facility database
- 10 Gbps General stream for email, web surfing, VOIP, video, diagnostic, monitoring.

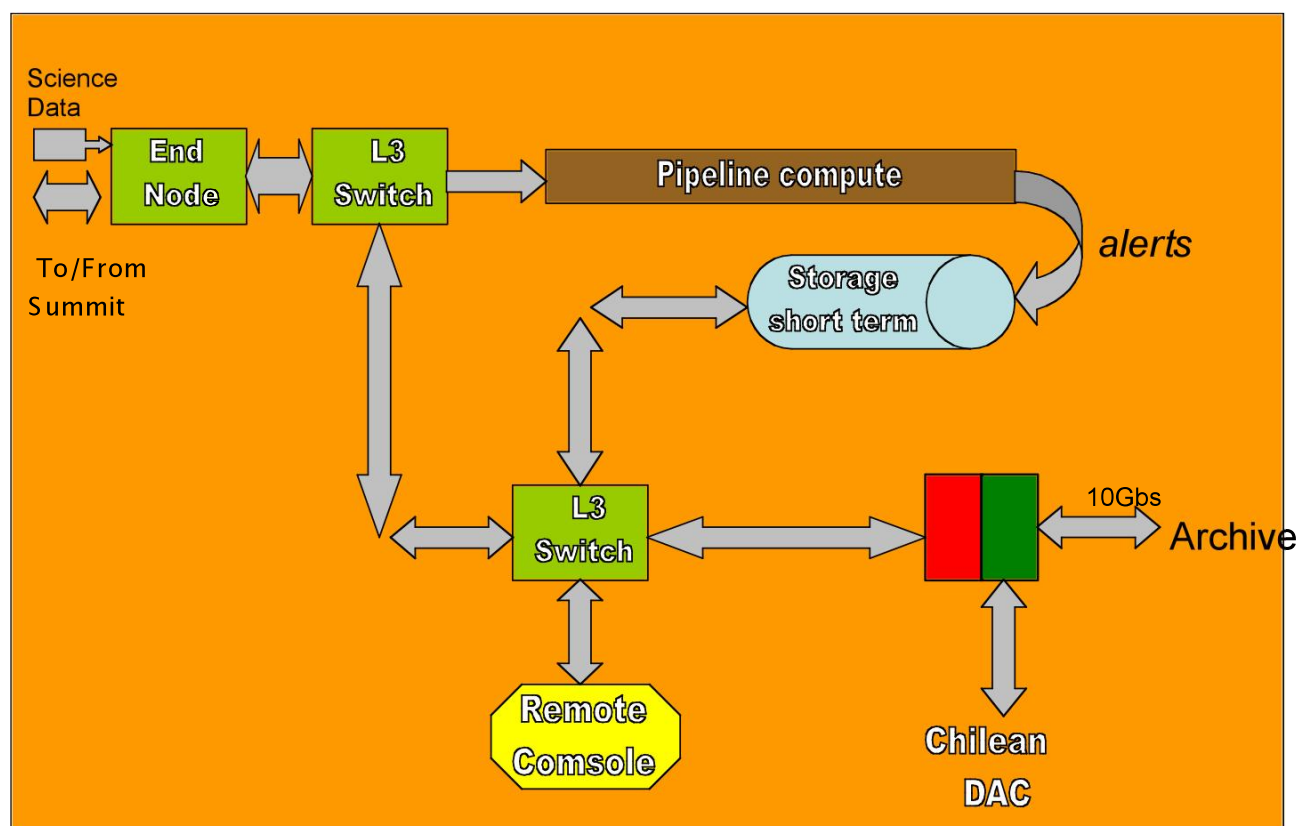
## Base Infrastructure

The Base Facility network is situated in La Serena at the AURA compound. Similar to the summit, an Ethernet meshed core-access architecture will be developed. Major functions performed at the base are the nightly pipeline for alerts, remote control room operations, a temporary two day store of data, and transmission infrastructure to the U.S. LSST will need a 25 TFLOPS machine in order to process images to produce alerts.

The pipeline cluster will consist of 3200 cores and will be inter-connected with a multi-port 10Gbps switch. The Science data will flow from the Summit through the pipeline compute into the short term storage cache of 150TB, and will also be transferred to the Archive Center. The raw image data will bifurcate in the Base Facility computer room to the Chilean Data Access Center logically located outside of the Base Facility firewall.

The bandwidth to the wide area network will be 10Gbps on which LSST will transfer 15TB per night to the Archive Center at NCSA. This bandwidth in Chile will support the National and Latin American Astronomical communities and public at large.

VLAN trunking will be employed across all switches between the Base Facility and Mountain Summit to maintain homogenous networks between Telescope, Camera, and Data Management traffic for security.

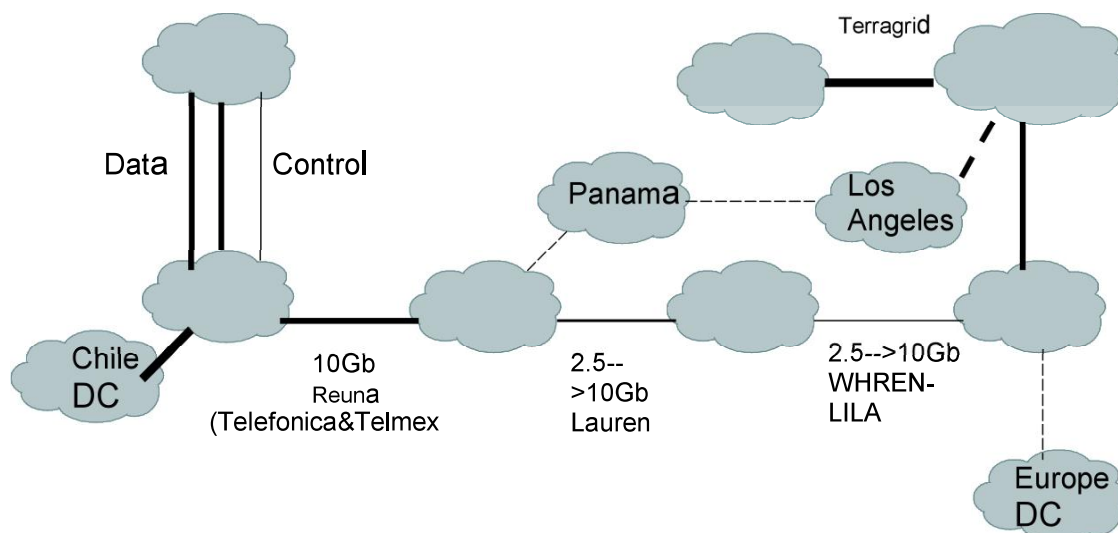


## International Network

Since 2003 \$10M (\$4M NSF) has been invested in U.S. – Latin America connectivity. The East side of South America will be at 10 Gbps by 2009, driven largely by high energy physics (HEP) and the Large Hadron Collider (LHC) project. Buenos Aires to Miami is now 2.5 Gbps. Brazil (RNP) is investing internally for 10 Gbps between Rio de Janeiro, Brasilia, and Sao Paolo and is formulating plans for upgrading the Brazil to U.S. links to 10 Gbps. On the West side,



Santiago to Miami is now 2 x 1 Gbps with a link from Panama across the Caribbean to Miami. The 1 Gbps Panama to San Diego spur is only Global Crossings fiber, and tends to be more expensive, but is also available.



Three providers (REUNA, LAUREN and WHREN) will be employed to reach the U.S from La Serena and the entire circuit will utilize DWDM technology.

Within Chile, REUNA (Chilean National Research and Education Networks) will provide a 10Gbps link from La Serena to Santiago. This link, which provides the circuit to Miami, will also be used extensively by the Chilean and Latin American communities to access the La Serena Data Access Center. The Latin American traffic will interact with Chile over CLARA, a South American network.

Two independent fiber paths will each be configured using independent paths. Each path will be provisioned as two STM-16 (5 Gbps net) capacity. While both fiber paths are operational a net of 10 gbps can be used. When there is a fiber cut, the other path is all that is available. This is the typical mechanism for “protected linear” circuits, but most commercial applications cannot survive throttling of the bandwidth, so in most cases one side is dormant until a cut, when the other side is used.

Inside of Chile, the three telecom operators were approached to offer solutions for this capacity. Two offered viable technical solutions that fit within an average monthly recurring cost of \$50,000 U.S. dollars with availability in 2012. Commercial letters of offer were made, insuring the creditability of the solution.

The link from Santiago to Sao Paulo will utilize the Global Crossing fiber ring operated by LAN NAUTILUS and provided by LAUREN. LSST has contracted an IRU with a private protected clear channel of 2.4 Gbps with access to burst to 10 Gbps during periods of catch up or abnormal transfer rates.

The Sao Paulo to Miami traffic will share the WHREN network of 10 Gbps provided by an NSF and Brazilian collaboration.

The operators of the two protected networks linking Chile to the U.S. were engaged to quote a corresponding amount of guaranteed and burst bandwidth. The network operators were asked to offer quotations for various solutions, the most response that fit both application needs and budget requirements was to provision two STM-16s protected from Santiago to Sao Paulo, and 10gb/s unprotected from Santiago to Miami. From Sao Paulo to Miami, the 2.5 gb/s of protected traffic would be transited via the NSF and ANSP supported LILA project, offering no-additional cost transit.

The unprotected circuit works differently in a ring configuration than a linear one. In a ring configuration a single cut results in all the protected traffic reversing course on to the unprotected bandwidth space. This means the likelihood for outages is greater than in a linear configuration, all other aspects being equal. In practice, no public data is available but the monitoring data suggest that there could be as much as 1/12 down time on the ring network. The down time is due to a recurring problem of thieves believing that there is copper in the fiber cable passing the Andes. This has become

compounded lately as they now target the repair trucks and tools of the staff dispatched to repair the fiber cuts. The desire for enough application bandwidth, and the expectation of potential disruptions due to ring cuts was a catalyst for the hybrid protected and unprotected capacity.

One vendor responded with the capacity required and did so for the budgeted cost of \$8.4 one time cost and \$710,000 annual operations and maintenance costs.

The interconnection between Miami and NCSA will be across FLR, NLR, and i-Wire provisioned as a 10 Gbps Ethernet solution. Network access fees, transponder and equipment costs are today, approximately \$40,000 a month (this is based on a similar project "AtlanticWave").

## Future Variables

While the cost of equipment may go down in the time period involved, access to network resources is likely to be more scarce. The costs may be significantly reduced, or may stay the same based on market condition changes between now and project roll-out.

A for-profit company linked to both the University of Chile, NTT Japan, and mining interests in Chile is considering the development of a high performance network back-bone providing access to lambdas to customers. They consider Chilean mining interests and LSST to be potential best first customers. If this venture is successful, it is unlikely that it will result in significant cost savings to LSST. The cost of capital to undertake such a project make it unlikely that with a small customer base that the cash flow needs would be the same or higher than the other commercial offerings. The advantage to LSST of any such project would be the proposition to be able to increase the amount of bandwidth available (in this scenario almost certainly two 10 Gbps paths for the same cost as 5 Gbps. Projects requiring significantly more bandwidth- in the 50 Gbps and higher range would realize significant savings, as commercial providers do not offer a standard product to meet those needs. LSST's needs of a few STM-16 fit into the suite of standard offerings.

Several projects are underway in Brazil to take advantage of TransitRail and lower cost commodity peering in the U.S. by purchasing leases on circuits on both the Atlantic and Pacific side. It is anticipated in 2007 a 1.2 Gbps Sao Paulo-Santiago-Tijuana-California link will be in place for 50% commodity and 50% research traffic. The value to LSST is that as more of these projects begin, and are able to leverage previous investments, it is easier to gain more bandwidth. Due to a monopoly on the legs interconnecting Santiago with Tijuana, Pacific capacity is very expensive. But through a consortium with WHREN it is likely that LSST will be able to have access to 5-10 Gbps path on the Pacific and well as the Atlantic side without additional expense.

Bandwidth requirements for astronomy, physics, and other sciences in Chile will also increase in the years as LSST becomes fully operational, which will drive an increase in availability. The non-LSST anticipated bandwidth requirements in Chile for astronomy alone are shown below.

Astronomy Program	Bandwidth Required (Gbps)	1 <sup>st</sup> Year Operations
CTIO (DECM Gemini, SOAR)	0.50	2010
ESO (VST, VISTA, E-ELT, etc.)	0.50	2010 - 2016
ALMA	0.15	2012
TOTAL	1.15	2010 - 2016

## Acknowledgements

LSST is a public-private partnership. The LSST design and development activity is supported by the National Science Foundation under Scientific Program Order No. 9 (AST-0551161) through Cooperative Agreement AST-0132798. Additional funding comes from private donations, in-kind support at Department of Energy laboratories and other LSSTC Institutional Members.

## Acronym Glossary

ALMA	Atacama Large Millimeter Array
ANSP	Academic Network at Sao Paulo
AURA	Association of Universities for Research in Astronomy
CCS	Camera Control System
CLARA	Latin American Collaboration for Advanced Networks
DAC	Data Access Center
DEC	Dark Energy Camera
DMS	Data Management System
E-ELT	European Extremely Large Telescope
ESO	European Southern Observatory
FLR	Florida Lambda Rail
Gbps	Gigabits per second
HEP	High Energy Physics
IRU	Indefeasible Right to Use
LAUREN	Latin American University Research and Education Networks
LHC	Large Hadron Collider
LILA	Links Interconnecting Latin America
LSST	Large Synoptic Survey Telescope
NCSA	National Center for Supercomputing Applications
NLR	National Lambda Rail
NSF	National Science Foundation
OCS	Observatory Control System
REUNA	Chilean National Research and Education Networks
STM	Synchronous Transport Module
TCS	Telescope Control System
TFLOPS	Trillion floating point operations per second
VISTA	Visible and Infrared Survey Telescope for Astronomy
VLAN	Virtual Local Area Network
VLT	Very Large Telescope
VST	VLT Survey Telescope
WCS	World Coordinate System
WHREN	Western Hemisphere Research and Education Networks